1. Pendahuluan

Hadis adalah segala sesuatu yang dikaitkan dengan Nabi Muhammad dalam bentuk perkataan, perbuatan dan taqrir (sikap diam) [1]. Dengan berkembangnya teknologi dan ilmu pengetahuan, kemajuan juga telah dicapai dalam pengembangan hadits. Buku-buku yang digunakan di atas berasal dari berbagai koleksi hadis yang dikemas dalam bentuk digital. Namun, kebanyakan hadis digital hanya menampilkan terjemahan hadis, tanpa penjelasan untuk setiap hadits, baik itu nasihat, larangan atau informasi. Hal ini memungkinkan umat Islam yang memahami dan mempelajari ilmu hadits salah mengartikan makna hadis. [2].

Pada tahun 2018, dilakukan penelitian [3] dibangun sebuah sistem klasifikasi topik multi-label dan single-label pada hadis *Bukhari* dengan menggunakan *Backpropogation Neural Network* dengan kategori kelas anjuran, larangan, dan informasi. Seleksi fitur yang digunakan adalah IG lalu metode untuk evaluasi menggunakan *Hamming Loss*. Penelitian tersebut mendapatkan akurasi terbaik diangka 88,42% untuk data hadis multi label dan 66,27% untuk data hadis berlabel tunggal [3].

Tahun 2017, pada penelitian [5], kombinasi *Multinomial* NB dan TFIDF digunakan untuk membangun sistem dapat mengelola teks yang terdiri dari 4 kategori, yaitu undangan, cuti, perintah, dan pinjam ruang. Hasil evaluasi sistem klasifikasi surat keluar didapatkan nilai Accuracy sebesar 89% [5].

Berdasarkan penelitian yang telah dibahas sebelumnya, kombinasi terhadap pengkategorian *corpus* dari hadis hadis masih kurang untuk terjamahan bahasa Indonesia. Penulis mengusulkan model klasifikasi *multi-label* hadis Bukhari dengan menggunakan metode *Information Gain* (IG) untuk *feature selection* dan metode klasifikasi *Multinomial Naïve Bayes* dengan kategori kelas anjuran, larangan, dan informasi. Metode *information gain* digunakan karena bekerja dengan memberikan nilai pada setiap fitur, dan hanya fitur terpilih yang terkait dengan kategori yang akan digunakan [3], sedangkan metode *Multinomial Naïve* dipilih karena cocok untuk untuk data yang memiliki kosakata yang banyak dan dapat mengurangi tingkat kesalahan [5].

.

Topik dan Batasannya

Pada penelitian penulis akan membangun model klasifikasi hadis Bukhari terjemahan bahasa Indonesia. Model yang dibangun menggunakan beberapa proses yaitu *perpocessing*, ektraksi fitur, seleksi fitur, dan klasifikasi *multinomial naïve bayes.* Dataset yang digunakan berjumlah 7008 dan berasal dari dataset terjemahan hadis Al Bukhari. Kelas yang terdapat pada dataset berjumlah 3 kelas, yaitu anjuran, larangan dan informasi.

Tujuan

Peneltian ini dilakukan untuk mengetahui *corpus* yang ada pada hadis dengan menggunakan beberapa skenario, yaitu pengaruh teknik *stemming* dan *stopword removal* pada *preprocessing*; nilai *treshold* pada penggunaan *Information Gain*; serta penggunaan data *balance* dan data *imbalance* pada klasifikasi *Multinomial Navive Bayes*.

Organisasi Tulisan

Bagian selanjutnya pada penelitian yaitu pembahasanan mengenai studi terkait pada peneleitian, penjelasan mengenai sistem yang dibangun, kemudian evaluasi hasil pengujian, pembahasan mengenai kesimpulan dari penelitian ini, dan terakhir pemabahasan saran untuk penelitian selanjutnya.